



**NetApp™**  
Go further, faster

## Success Stories

# Genome Sciences Centre Keeps Pace with Storage Growth and Network Performance



### KEY HIGHLIGHTS

#### Industry

Biotechnology

#### The challenge

Scale network to tens of millions of files, increase reliability, and ensure recovery

#### The solution

NetApp® storage systems with NetApp SnapMirror® and NetApp Operations Manager

#### Benefits

- Saved hundreds of thousands of dollars as a result of no failures
- Increased storage capacity by over 1,000%
- Slashed time needed to back up files by over 90%
- Doubled the number of network nodes in three years, from 200 to 400
- Gained efficiencies, saving one headcount

### CUSTOMER PROFILE

The Michael Smith Genome Sciences Centre (GSC) carries out genomic research and bioinformatics using one of the most advanced computing platforms in the world. The center operates at the BC Cancer Agency and employs over 200 scientists and researchers ([www.bcgsc.ca](http://www.bcgsc.ca)).

### THE CHALLENGE

#### Turning genes off and on

Researchers at the Michael Smith Genome Sciences Centre (GSC) in Vancouver, British Columbia, want to tackle cancer at its source. Operating under the auspices of the BC Cancer Agency, the 200 scientists at GSC perform experimental genomics using one of the most innovative bioinformatics computing platforms in the world, built on a high-performance Linux® cluster.

Thanks to the genomics work at GSC and elsewhere, scientists are learning through computer analysis how to “turn off” the genes that drive cancer growth and “turn on” our body’s best defenses against cancer and other diseases. Not surprisingly, this requires massive amounts of computing power and storage. Five years ago, rapid advances in genomic technology led to

the data stored at the lab zooming from 1TB to 18TB. Meanwhile, the GSC commodity-hardware Linux cluster topped the 200-node mark as the network grew in complexity.

“We were bursting at the seams in terms of Linux file server capacity,” says Greg Stazyk, systems coordinator at GSC. “Sequencing production processes generate tens of millions of files. In fact, a single volume can contain nearly 30 million files. We had absolutely reached the performance and manageability limits of the existing file server structure.” A number of disruptive file server crashes underscored the urgency of finding a more robust solution.

#### Dealing with a data explosion

Server crashes and storage manageability problems disappeared after the center acquired its first NetApp FAS unified storage solution in 2003. Today, NetApp FAS systems are at the heart of a fast-growing unified storage solution with over 800 nodes and multiple storage arrays. The hardware and software from NetApp provide network-attached storage (NAS) for all lab data, including production pipelines for genome mapping and DNA sequencing.

“One of the things that makes NetApp great is its support of all of the industry-standard protocols we are using. Our regular network guys can easily do all of the maintenance and provisioning operations themselves with minimal additional training.”

**Greg Stazyk**  
Systems Coordinator, GSC

Where older-generation sequencers typically generate 2GB to 3GB of data per week, Stazyk says three new sequencers acquired by the lab in 2007 crank out an order of magnitude more—nearly 1TB of data each week. As a result, data demands continue to skyrocket. “Because so much of the work here is cutting edge, our scientists typically keep everything their research generates and are doing more with the data than ever before.”

Solution cost and complexity remain important considerations for a lab that is funded largely through grants. As is the case at most research organizations, “It’s easier to find budget for science than budget to run an IT group,” says Stazyk. “With limited headcount to support our IT infrastructure, we always strive to deploy highly flexible and manageable technologies such as the NetApp FAS systems we now have in place.”

#### **THE SOLUTION**

##### **Adding to the infrastructure**

NetApp’s solutions continue to keep pace with the organization’s voracious demands for storage. A recent addition is a NetApp FAS6070, a large enterprise-class array system noted for its flexibility and support

of a wide range of protocols in use at the center, including NFS, HTTP, and iSCSI. “One of the things that makes NetApp great is its support of all of the industry-standard protocols we are using,” says Stazyk. “Our regular network guys can easily do all of the maintenance and provisioning operations themselves with minimal additional training.”

The FAS6070 storage system complements two midrange NetApp FAS3050 systems that have been upgraded with more storage over the past two years. A slightly older NetApp NearStore® system is currently being converted from primary storage duties to disk-to-disk backup, archiving, and secondary content storage. The total amount of online storage available is approaching 300TB, with the potential to go much higher. “The NetApp solution we currently have in place will take us well into the petabyte range,” says Stazyk, “and at our current rate of growth we’re going to need it.”

NetApp’s solution keeps production pipelines operating interruption free while it also supports more than 200 science and office workers running an array of Linux desktop and Windows® applications. “We move a lot of data around, especially during

heavy analysis on the Linux cluster,” says Stazyk. “Performance has been excellent, even when the network is being pounded.”

The data load on the network recently picked up substantially when GSC upgraded its high-speed network from 1 Gigabit Ethernet to 10 Gigabit Ethernet speeds, but thanks to careful planning, overall performance was not affected. “We’ve had to improve all of our other systems to keep pace with the rate of storage,” says Stazyk. For example, the GSC servers run the latest dual-core Opteron processors, each equipped with 16GB of RAM.

#### **BUSINESS BENEFITS**

##### **A reliable network**

Since the deployment of the solution from NetApp, there have been no process-stopping failures, a benefit that Stazyk estimates has saved hundreds of thousands of dollars. “When a failure occurred in the old file server structure, there was a risk that we could have to restart a weeks-long process. Biologist productivity was negatively affected, and the potential existed to lose large quantities of the costly chemical and reaction agents used in the sequencing processes.” Stazyk says that today’s

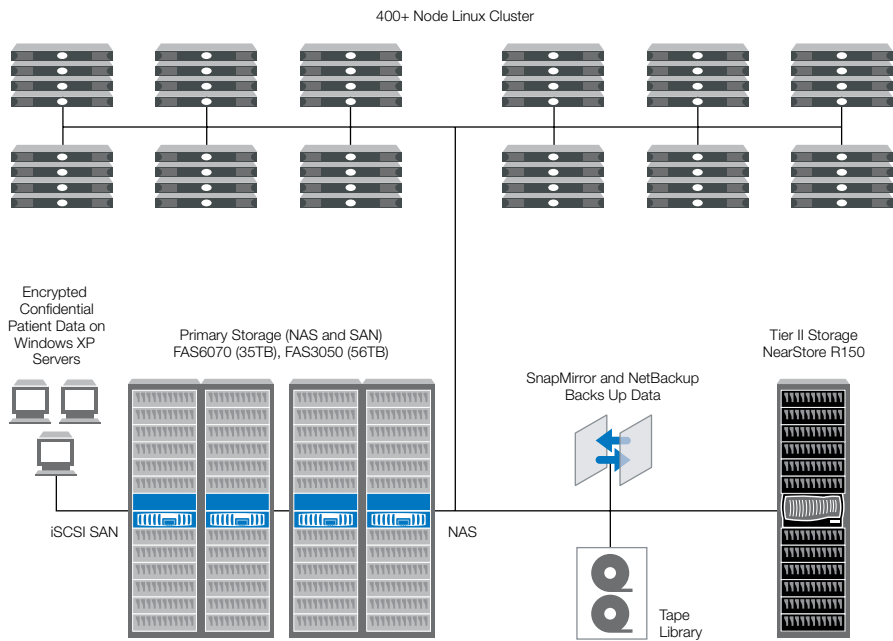


Figure 1) An enterprise-class NetApp FAS6070 array provides primary NAS storage using NFS with iSCSI support included for common office files. Additional primary storage is handled by two midrange NetApp FAS3050 systems, with secondary storage provided by a NetApp NearStore 150 and tape.

solution is much more reliable, with the added benefit of NetApp Snapshot™ technology for near-instantaneous data recovery.

NetApp RAID-DP®, which allows dual striping for recovery from a multidisk failure in a single volume, adds another important level of protection.

For data backups, Stazyk’s IT team relies on a mix of NetApp SnapMirror software, the NetApp NDMP backup protocol, and Veritas™ NetBackup™ from Symantec. “We’ve found that using SnapMirror is the best way for us to move a lot of data around, which we do quite frequently,” says Stazyk. “Before SnapMirror, it would take us nearly 12 hours to copy 5TB of data. But with SnapMirror, once we do the initial copy, the only data sent over the network is the differentials. That process usually takes less than an hour and has been very reliable.”

Adding to the reliability and cost effectiveness is the tight integration between the NetApp storage virtualization solutions and the server virtualization software from VMware. “We can use Snapshot copies to create images of the virtual machines and move those environments easily from one

environment to the next,” says Stazyk. “Just being able to serve virtual machines off of NFS is a big advantage for us. With NetApp and VMware® together, we’ve been able to retire a number of old servers and increase reliability and network speed all at the same time.”

**Easy to provision and expand**

NetApp on-demand storage scalability ensures that production does not stop for expansion. “It’s important that we don’t run out of capacity in the middle of a large job,” says Stazyk. Fortunately, NetApp Operations Manager automatically identifies when a process under way is close to capacity. When a recent process that had been running for nearly two weeks came close to running out of space, Stazyk says it took just 20 minutes to bring an additional 2TB online. “We added the shelf while the NetApp system was live and pushing over 30,000 NFS ops per second,” he says. “If we had stopped to add capacity, we would have lost two weeks of computation.”

**A secure SAN**

In addition to using systems from NetApp for scientific storage, GSC also runs a separate NetApp iSCSI SAN for secure storage of office-type documents. This network helps

the center protect sensitive patient data and comply with Canadian information privacy laws. The Personal Information Protection and Electronic Documents Act (PIPEDA) specifies the types of personal data that can be collected, how it is stored, and precautions that must be taken to protect the data.

“PIPEDA strictly regulates the storage of any data that can be tied back to an individual anywhere in the country,” says Stazyk. “Storing Windows XP encrypted patient information on the NetApp iSCSI SAN allows us to easily segregate and automatically back up encrypted data.”

**Counting the benefits**

Stazyk and his team are well aware of the benefits that NetApp brings to their data center. For example, “I tend to think that reliability is at an all-time high,” says Stazyk. “The servers don’t go down anymore, and everything just runs as it’s supposed to. And if any data ever gets accidentally deleted, we have NetApp Snapshot copies to always recover.”

By making the right hardware and software purchases, Stazyk says he’s been able to forego hiring additional people in the

“With limited headcount to support our IT infrastructure, we always strive to deploy highly flexible and manageable technologies like the NetApp FAS systems we now have in place. In essence we’ve saved one headcount by buying good capital equipment. We’re doing a lot more for a lot less.”

Greg Stazyk  
Systems Coordinator, GSC

systems group. “In essence, we’ve saved one headcount by buying good capital equipment. We’re doing a lot more for a lot less.”

Stazyk’s team once spent a lot of time dealing with file server and NFS issues and creating islands of storage for various purposes. NetApp’s simple, unified structure makes most of those issues a thing of the past. Perhaps best of all, the network structure is equally simple from an end-user perspective. Biologists, understandably, don’t want to waste time dealing with complex data access patterns. Rather, says Stazyk, “They want to do science. That’s what NetApp’s storage solution enables.”

#### SOLUTION COMPONENTS

##### NetApp Products

NetApp FAS6070

NetApp 3050

NetApp NearStore R150

NetApp SnapMirror

NetApp Operations Manager

NAS and iSCSI protocols

##### Partner Products

VMware server virtualization software

Veritas NetBackup from Symantec

##### Environment

Applications: Genomic sequencing software and various research applications; Microsoft® Office

Server platform: Generic servers with 16GB of RAM running Linux and VMware

NetApp creates innovative storage and data management solutions that accelerate business breakthroughs and deliver outstanding cost efficiency. Discover our passion for helping companies around the world go further, faster at [www.netapp.com](http://www.netapp.com).

© 2008 NetApp. All rights reserved. Specifications are subject to change without notice. NetApp, the NetApp logo, Go further, faster, NearStore, RAID-DP, SnapMirror, and Snapshot are trademarks or registered trademarks of NetApp, Inc. in the United States and/or other countries. Microsoft and Windows are registered trademarks of Microsoft Corporation. Linux is a registered trademark of Linus Torvalds. Veritas and NetBackup are trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. VMware is a registered trademark of VMware, Inc. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such. CSS-4091-0508



[www.netapp.com](http://www.netapp.com)